A Modified Nonparametric Prewhitened Covariance Estimator

Masayuki Hirukawa
Concordia University
Montréal, Québec, Canada

October 21, 2005

1 Motivations

- Prewhitening (Andrews and Monahan, 1992) often yields poor empirical sizes of the test statistic, as indicated in Monte Carlo simulations in Hirukawa (2005).

→ An alternative (possibly nonparametric) bias correction method for HAC estimation?

- The GMM bootstrap requires a HAC estimator with a higher-order kernel for asymptotic refinements (Hall and Horowitz, 1996; Andrews, 2002, 2004; Inoue and Shintani, 2005).

→ A HAC estimator that has the same order of magnitude in the bias term as with a higher-order kernel but maintains positive semidefiniteness in finite samples?

2 Multiplicative Bias Correction

A solution would be (a fully modified version of) the **nonparametric prewhitened (NPW) covariance estimator** in Xiao and Linton (2002, abbreviated as "XL"). This estimator applies a *multiplicative bias correction*. This method is based on the following simple idea. Suppose that an estimator is known to have the bias $b_n(\theta)$ such that

$$E\left(\hat{\theta}\right) = \theta \cdot b_n(\theta),$$

where

$$b_n(\theta) = 1 + o(1),$$

a multiplicatively bias-corrected estimator $\hat{\theta}^M$ is given by

$$\hat{\theta}^M = \frac{\hat{\theta}}{b_n\left(\hat{\theta}\right)}.$$

3 Related Literature

XL's NPW spectral matrix estimator (and the long-run variance estimator as a special case) belongs to the same family as the following estimators. Each estimator is $n^{4/9}$-consistent (faster than the usual rate of $n^{2/5}$!) when a second-order kernel is employed. Nonetheless, it maintains positivity in finite samples.

Nonparametric regression: Linton and Nielsen (1994).

Probability density estimation: Jones, Linton and Nielsen (1995).

Hazard estimation: Nielsen (1998), Nielsen and Tanggaard (2001).

4 Contributions

- XL's definition of the NPW estimator does not necessarily lead to a positive semidefinite estimate.

→ *Proposed the modified NPW estimator.*

- The leading variance term of XL's NPW estimator does not involve the roughness of the spectral window obtained by twicing (Stuetzle and Mittal, 1979).

→ *Derived an asymptotic expansion of the estimator and showed that the variance term indeed involved the roughness.*

- For a given frequency, XL's definition of the MSE does not necessarily yield a nonnegative squared bias or variance term.

→ *Provided a sensible definition of the MSE.*

- No asymptotic expansion to the NPW iteration is available in XL.

→ *Derived an asymptotic expansion of the iterated NPW estimator.*

- XL's Monte Carlo simulations are invalid.

→ *Conducted Monte Carlo simulations based on the modified NPW estimator.*

5 A Modified NPW Estimator

The (bias-uncorrected) spectral matrix estimator of a stationary vector time series $\{\mathbf{x}_t\} \in \mathbb{R}^d$ at the frequency $\omega \in (-\pi, \pi)$ is given by

$$\hat{f}_{xx}(\omega) = \frac{2\pi}{T} \sum_{\lambda_j \in B(\omega)} K_M\left(\lambda_j - \omega\right) I_{xx}\left(\lambda_j\right),$$

where $\lambda_j = 2\pi j/T$ is the fundamental frequency, $B(\omega) = \left\{\lambda_j \middle| \omega - \pi < \lambda_j < \omega + \pi\right\}$ is the frequency band centered at the frequency ω, $K_M(\cdot)$ is the amplitude window (Parzen, 1963) with bandwidth $M \in \mathbb{R}_+$ implied by the spectral window $K(\cdot)$, and $I_{xx}\left(\lambda_j\right)$ is the periodogram at the frequency λ_j. Then, the modified NPW spectral matrix estimator becomes

$$\begin{aligned} & \tilde{f}_{xx}(\omega) \\ = \; & \hat{f}_{xx}^{1/2}(\omega)\, \tilde{\alpha}(\omega)\, \hat{f}_{xx}^{1/2}(\omega) \\ = \; & \hat{f}_{xx}^{1/2}(\omega) \left\{ \sum_{\lambda_j \in B(\omega)} K_M\left(\lambda_j - \omega\right) \right. \\ & \left. \times \frac{2\pi}{T} \hat{f}_{xx}^{-1/2}\left(\lambda_j\right) I_{xx}\left(\lambda_j\right) \hat{f}_{xx}^{-1/2}\left(\lambda_j\right) \right\} \hat{f}_{xx}^{1/2}(\omega), \end{aligned}$$

where $\tilde{\alpha}(\omega)$ serves as the bias correction term.

Similarly, by $\Omega = 2\pi f_{xx}(0)$, the (bias-uncorrected) HAC estimator is expressed as a smoothed periodogram

$$\hat{\Omega} = \frac{4\pi^2}{T} \sum_{\lambda_j \in B(0)} K_M\left(\lambda_j\right) I_{xx}\left(\lambda_j\right).$$

Then, the modified NPW covariance matrix estimator becomes

$$\begin{aligned}
& \tilde{\Omega} \\
= \; & \hat{\Omega}^{1/2} \tilde{\alpha}(0) \hat{\Omega}^{1/2} \\
= \; & \hat{\Omega}^{1/2} \left\{ \sum_{\lambda_j \in B(0)} K_M\left(\lambda_j - \omega\right) \right. \\
& \left. \times \frac{2\pi}{T} \hat{f}_{xx}^{-1/2}\left(\lambda_j\right) I_{xx}\left(\lambda_j\right) \hat{f}_{xx}^{-1/2}\left(\lambda_j\right) \right\} \hat{\Omega}^{1/2}.
\end{aligned}$$

When the spectral window $K(\cdot)$ is a second-order one, $\tilde{f}_{xx}(\omega)$ is Hermitian and positive definite in finite samples by the "sandwich form."

6 Asymptotic Expansion

Under the assumption of sufficient smoothness in the spectral density, $\tilde{f}_{xx}(\omega)$ has an asymptotic expansion

$$\begin{aligned}\tilde{f}_{xx}(\omega) &= f_{xx}(\omega) + \tilde{\mathcal{B}}(\omega) + \tilde{\mathcal{V}}(\omega) \\ &\quad + o_p\left(M^{-4} + \left(\frac{T}{M}\right)^{-1/2}\right).\end{aligned}$$

The leading bias term $\tilde{\mathcal{B}}(\omega)$ is approximated by

$$M^4 \tilde{\mathcal{B}}(\omega) \simeq -k_2^2 \Psi(\omega),$$

where

$$\begin{aligned}k_2 &= \lim_{x \to 0} \frac{1 - k(x)}{|x|^2} \in (0, \infty), \\ k(x) &= \int_{-\infty}^{\infty} K(\theta) e^{ix\theta} d\theta, \\ \Psi(\omega) &= f_{xx}^{1/2}(\omega) \Phi''(\omega) f_{xx}^{1/2}(\omega), \\ \Phi(\omega) &= f_{xx}^{-1/2}(\omega) f''_{xx}(\omega) f_{xx}^{-1/2}(\omega).\end{aligned}$$

The leading variance term $\tilde{\mathcal{V}}(\omega)$ is approximated by

$$\frac{T}{M} Var\left(\text{vec}\left(\tilde{\mathcal{V}}(\omega)\right)\right)$$

$$\simeq \begin{cases} 2\pi \int_{-\infty}^{\infty} T_K^2(\theta)\, d\theta f_{xx}(\omega) \otimes f_{xx}(\omega)^\top \\ \text{for } \omega \neq 0 \\ 2\pi \int_{-\infty}^{\infty} T_K^2(\theta)\, d\theta \left(I_{d^2} + K_{dd}\right) f_{xx}(0) \otimes f_{xx}(0) \\ \text{for } \omega = 0 \end{cases}$$

$$\simeq \begin{cases} \int_{-\infty}^{\infty} t_k^2(x)\, dx f_{xx}(\omega) \otimes f_{xx}(\omega)^\top \\ \text{for } \omega \neq 0 \\ \int_{-\infty}^{\infty} t_k^2(x)\, dx \left(I_{d^2} + K_{dd}\right) f_{xx}(0) \otimes f_{xx}(0) \\ \text{for } \omega = 0 \end{cases},$$

where $\circ$ denotes convolution,

$$T_K(\theta) = 2K(\theta) - K \circ K(\theta)$$

is the fourth-order spectral window obtained by twicing (Stuetzle and Mittal, 1979), and

$$t_k(x) = \int_{-\infty}^{\infty} T_K(\theta)\, e^{ix\theta} d\theta$$

is the kernel (or lag window) corresponding to the spectral window $T_K(\theta)$.

7 Definition of MSE

For each fixed ω, (i) $ABias \in \mathbb{R}$ or $(ABias)^2 \in \mathbb{R}_+$, and (ii) $AVar \in \mathbb{R}_+$ are the requirements for a sensible AMSE-optimal bandwidth choice. Since the spectral matrix estimator is complex-valued in general, the MSE in XL

$$\begin{aligned} & MSE\left(\tilde{f}_{xx}(\omega); f_{xx}(\omega)\right) \\ = \; & E\Big\{\mathrm{vec}\left(\tilde{f}_{xx}(\omega) - f_{xx}(\omega)\right)^{\top} \\ & W\,\mathrm{vec}\left(\tilde{f}_{xx}(\omega) - f_{xx}(\omega)\right)\Big\} \end{aligned}$$

does not necessarily yield a nonnegative squared bias or variance term. Then, for some weighting vector $v \in \mathbb{R}^d$, define

$$\begin{aligned} & MSE\left(\tilde{f}_{xx}(\omega); f_{xx}(\omega)\right) \\ = \; & E\left\{v^{\top}\left(\tilde{f}_{xx}(\omega) - f_{xx}(\omega)\right)v\right\}^2. \end{aligned}$$

This weighting scheme cancels the imaginary parts of the (i,j) and (j,i) entries in the bias and variance terms.

8 AMSE-Optimal Bandwidth Choice

$$
\begin{aligned}
& MSE\left(\tilde{f}_{xx}(\omega); f_{xx}(\omega)\right) \\
\simeq\ & \frac{k_2^4\left(v^\top \boldsymbol{\Psi}(\omega)\, v\right)^2}{M^8} \\
& + \frac{M}{T}\left(1 + \mathbf{1}(\omega = 0)\right)\left(v^\top f_{xx}(\omega)\, v\right)^2 \int_{-\infty}^{\infty} t_k^2(x)\, dx,
\end{aligned}
$$

where $ABias \in \mathbb{R}$ and $AVar \in \mathbb{R}_+$ are guaranteed. The bandwidth that minimizes the AMSE is

$$
\begin{aligned}
& M_{opt}(\omega) \\
=\ & \left\{\frac{8k_2^4\left(v^\top \boldsymbol{\Psi}(\omega)\, v\right)^2}{\left(1 + \mathbf{1}(\omega = 0)\right)\left(v^\top f_{xx}(\omega)\, v\right)^2 \int_{-\infty}^{\infty} t_k^2(x)\, dx}\right\}^{1/9} \\
& \times T^{1/9}.
\end{aligned}
$$

The AMSE-optimal bandwidth is $O\left(T^{1/9}\right)$. As a result,

$$
MSE_{opt}\left(\tilde{f}_{xx}(\omega); f_{xx}(\omega)\right) = O\left(T^{-8/9}\right),
$$

which is usually achieved under a fourth-order kernel.

Iteration of NPW

Define n times iterated NPW spectral matrix estimator as

$$\begin{aligned}
& \tilde{f}_{xx,n}(\omega) \\
= \; & \tilde{f}^{1/2}_{xx,n-1}(\omega)\,\tilde{\alpha}_n(\omega)\,\tilde{f}^{1/2}_{xx,n-1}(\omega) \\
= \; & \tilde{f}^{1/2}_{xx,n-1}(\omega)\left\{\sum_{\lambda_j\in B(\omega)} K_M\left(\lambda_j-\omega\right)\right. \\
& \left.\times\frac{2\pi}{T}\tilde{f}^{-1/2}_{xx,n-1}\left(\lambda_j\right) I_{xx}\left(\lambda_j\right)\tilde{f}^{-1/2}_{xx,n-1}\left(\lambda_j\right)\right\}\tilde{f}^{1/2}_{xx,n-1}(\omega),
\end{aligned}$$

where $\tilde{f}_{xx,0}(\omega) = \hat{f}_{xx}(\omega)$. Under the assumption of sufficient smoothness in the spectral density, $\tilde{f}_{xx,n}(\omega)$ has an asymptotic expansion

$$\begin{aligned}
\tilde{f}_{xx,n}(\omega) \; = \; & f_{xx}(\omega) + \tilde{\mathcal{B}}_n(\omega) + \tilde{\mathcal{V}}_n(\omega) \\
& + o_p\left(M^{-(2+2n)} + \left(\frac{T}{M}\right)^{-1/2}\right).
\end{aligned}$$

The leading bias term $\tilde{\mathcal{B}}_n(\omega)$ is approximated by

$$M^{2+2n}\tilde{\mathcal{B}}_n(\omega) \simeq (-1)^n\, k_2^{n+1}\Psi_n(\omega),$$

where

$$\begin{aligned}\Psi_n(\omega) &= f_{xx}^{1/2}(\omega)\,\Phi_n''(\omega)\,f_{xx}^{1/2}(\omega),\\ \Phi_n(\omega) &= f_{xx}^{-1/2}(\omega)\,\Psi_{n-1}(\omega)\,f_{xx}^{-1/2}(\omega),\\ \Psi_0(\omega) &= f_{xx}''(\omega).\end{aligned}$$

The leading variance term $\tilde{\mathcal{V}}_n(\omega)$ is approximated by

$$\begin{aligned}&\frac{T}{M}Var\left(\mathrm{vec}\left(\tilde{\mathcal{V}}_n(\omega)\right)\right)\\ \simeq&\begin{cases}2\pi\int_{-\infty}^{\infty}K_n^2(\theta)\,d\theta f_{xx}(\omega)\otimes f_{xx}(\omega)^\top\\ \text{for } \omega\neq 0\\ 2\pi\int_{-\infty}^{\infty}K_n^2(\theta)\,d\theta\left(I_{d^2}+K_{dd}\right)f_{xx}(0)\otimes f_{xx}(0)\\ \text{for } \omega=0\end{cases}\\ \simeq&\begin{cases}\int_{-\infty}^{\infty}k_n^2(x)\,dx f_{xx}(\omega)\otimes f_{xx}(\omega)^\top\\ \text{for } \omega\neq 0\\ \int_{-\infty}^{\infty}k_n^2(x)\,dx\left(I_{d^2}+K_{dd}\right)f_{xx}(0)\otimes f_{xx}(0)\\ \text{for } \omega=0\end{cases},\end{aligned}$$

where

$$\begin{aligned}K_n(\theta) &= K_0(\theta)+K_{n-1}(\theta)-K_0\circ K_{n-1}(\theta),\\ K_0(\theta) &= K(\theta),\\ k_n(x) &= \int_{-\infty}^{\infty}K_n(\theta)\,e^{ix\theta}d\theta.\end{aligned}$$

9 Monte Carlo Experiments

Purpose: Comparing the accuracy of estimates of the long-run variance (LRV) $\Omega = 2\pi f_{xx}(0)$ implied by several alternative spectral matrix estimators.

DGPs: ARMA(1,1), MA(2), and AR(2).

Sample size: 256.

Number of replications: 5,000.

Spectral estimators: The NPW estimator (*NPW*); the bias-uncorrected estimator (*BUC*); the AR(1)-prewhitened estimator (*AM*); and the trapezoidal lag-window estimator in Politis and Romano (1995) (*PR*).

Table 1: RMSEs and Biases of LRV Estimates for ARMA(1,1) Models

$$x_t = \rho x_{t-1} + \epsilon_t + \psi \epsilon_{t-1},\ \epsilon_t \overset{iid}{\sim} N(0,1).$$

ρ	ψ	Ω	NPW	BUC	AM	$PR1$
0.8	0.0	25.0000	10.6541	9.8213	9.8535	9.4755
			(-5.0036)	(-5.9529)	(-0.5427)	(-7.9805)
0.5	0.0	4.0000	1.0470	1.0022	0.9429	1.0578
			(-0.3574)	(-0.5312)	(-0.0249)	(-0.2768)
-0.5	0.0	0.4444	0.0745	0.0792	0.0543	0.1270
			(-0.0269)	(0.0359)	(-0.0020)	(-0.0091)
-0.8	0.0	0.3086	0.0661	0.0591	0.0384	0.1199
			(-0.0451)	(0.0290)	(-0.0022)	(0.0730)

Note: Numbers in the first and second rows of a given DGP are the RMSE and the bias (in parenthesis), respectively.

Table 1 (cont'd)

$$x_t = \rho x_{t-1} + \epsilon_t + \psi \epsilon_{t-1},\ \epsilon_t \overset{iid}{\sim} N(0,1).$$

ρ	ψ	Ω	NPW	BUC	AM	$PR1$
0.0	0.8	3.2400	0.4981	0.5664	1.5129	0.9098
			(-0.2725)	(-0.3016)	(1.2093)	(-0.1552)
0.0	0.5	2.2500	0.3389	0.3838	0.8794	0.6374
			(-0.1663)	(-0.1956)	(0.6755)	(-0.1071)
0.0	-0.5	0.2500	0.0626	0.0667	0.2086	0.0811
			(0.0056)	(0.0224)	(0.1985)	(-0.0088)
0.0	-0.8	0.0400	0.0196	0.0208	0.2954	0.0370
			(0.0042)	(0.0087)	(0.2903)	(0.0076)

Table 1 (cont'd)

$$x_t = \rho x_{t-1} + \epsilon_t + \psi \epsilon_{t-1},\ \epsilon_t \overset{iid}{\sim} N(0,1).$$

ρ	ψ	Ω	NPW	BUC	AM	$PR1$
0.5	0.8	12.9600	3.5759	3.4543	11.4281	3.5816
			(-1.1764)	(-1.7770)	(9.2036)	(-0.9033)
0.5	0.5	9.0000	2.4473	2.3709	6.9767	2.4537
			(-0.8460)	(-1.2583)	(5.5619)	(-0.6569)
0.5	-0.8	0.1600	0.0731	0.0748	0.5426	0.0930
			(0.0104)	(0.0283)	(0.5325)	(0.0375)
0.2	0.2	2.2500	0.4278	0.4340	0.5945	0.6274
			(-0.1351)	(-0.2231)	(0.3090)	(-0.1119)
-0.2	-0.2	0.4444	0.0843	0.0879	0.0951	0.1311
			(0.0467)	(0.0335)	(0.0686)	(-0.0179)
-0.5	0.8	1.4400	0.1788	0.2185	0.4047	0.4088
			(-0.0090)	(-0.1705)	(0.2891)	(-0.0775)
-0.5	-0.5	0.1111	0.0281	0.0328	0.0940	0.0559
			(0.0052)	(0.0137)	(0.0894)	(0.0207)
-0.5	-0.8	0.0178	0.0137	0.0147	0.1307	0.0646
			(0.0057)	(0.0078)	(0.1288)	(0.0420)

Figure 1: Spectral Density Estimates for Selected DGPs

MA(2) with $(\psi_1, \psi_2) = (0.0, 0.9)$

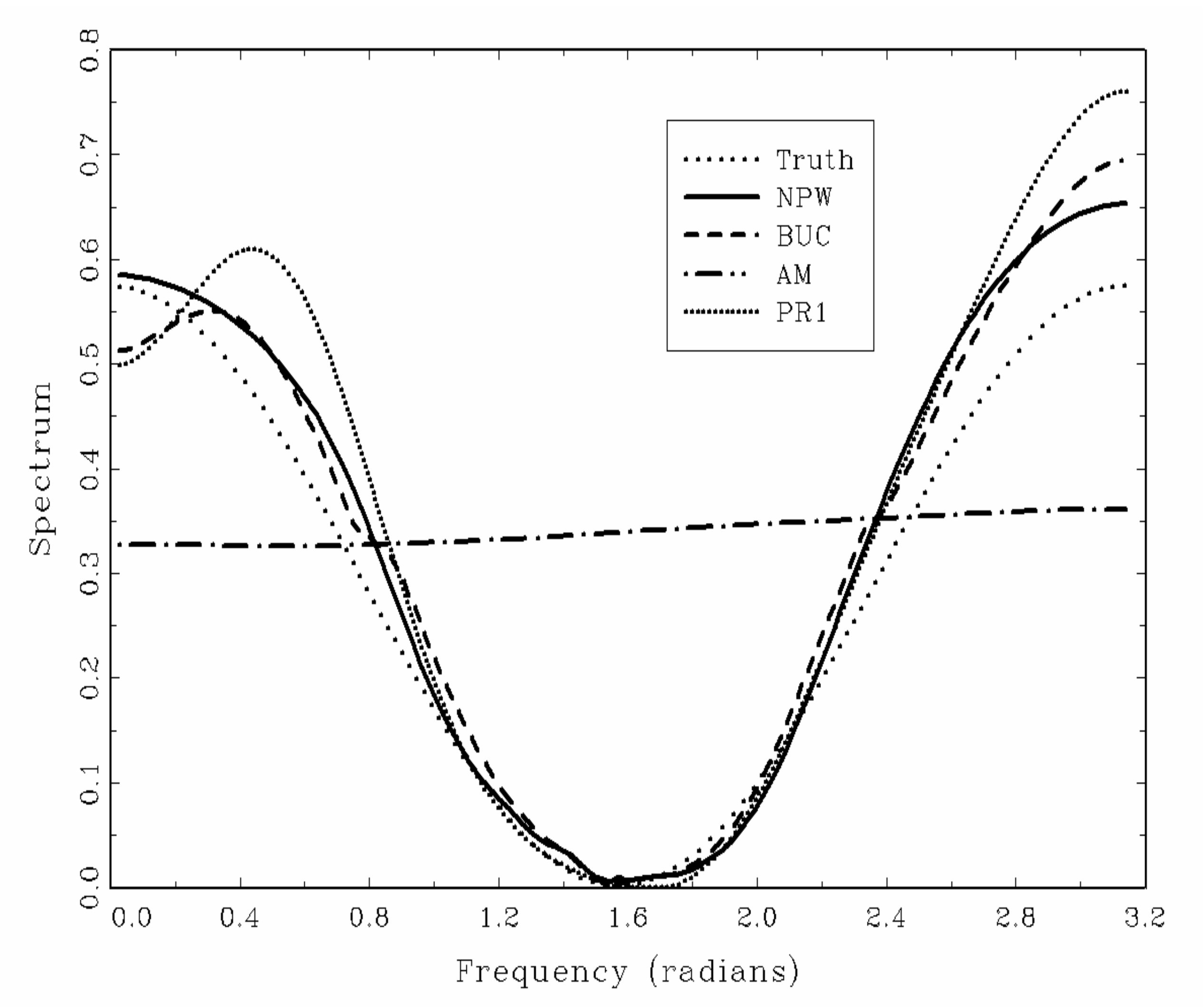
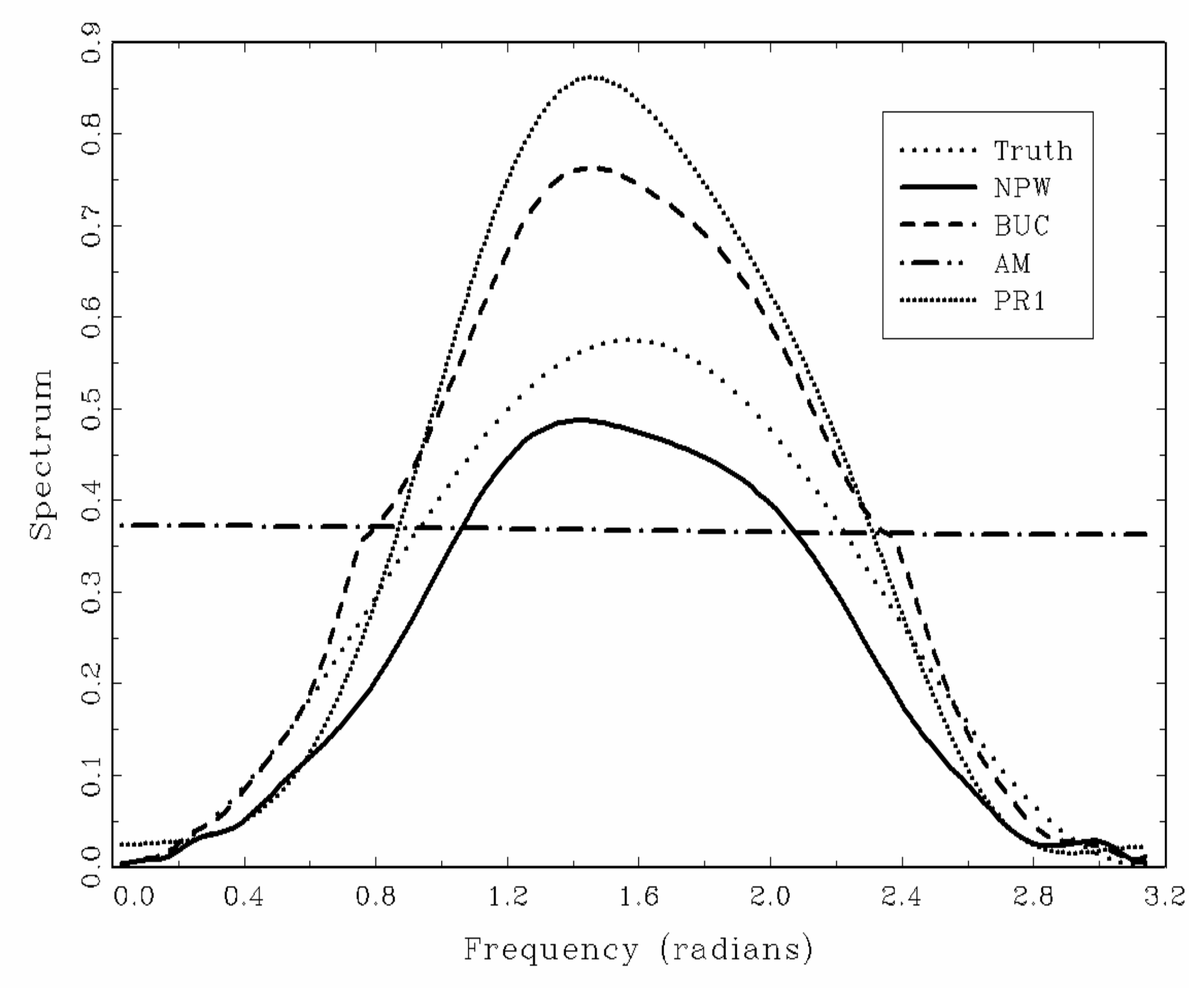


Figure 1 (cont'd)

MA(2) with $(\psi_1, \psi_2) = (0.0, -0.9)$



10 Conclusion

- The modified NPW estimator establishes the convergence rate of $O\left(T^{-8/9}\right)$ in MSE with the second-order spectral window, and it yields a Hermitian and positive semidefinite estimate.

- NPW can be iterated further to attain the convergence rate arbitrarily close to the parametric one.

- Monte Carlo results indicate that for a wide variety of DGPs the NPW estimator reduces the bias without inflating the variance.

- Research extensions (ongoing) are (i) to derive a reliable implementation method of the bandwidth, and (ii) to install the NPW covariance estimator into the GMM bootstrap.